The Computing and Data Grid Approach: Infrastructure for Distributed Science Applications

William E. Johnston[a]

Lawrence Berkeley National Laboratory[b] *and NASA Ames Research Center*[c]

Abstract

With the advent of Grids — infrastructure for using and managing widely distributed computing and data resources in the science environment — there is now an opportunity to provide a standard, large-scale, computing, data, instrument, and collaboration environment for science that spans many different projects and provides the required infrastructure and services in a relatively uniform and supportable way. Grid technology has evolved over the past several years to provide the services and infrastructure needed for building "virtual" systems and organizations. We argue that Grid technology provides an excellent basis for the creation of the integrated environments that can combine the resources needed to support the large-scale science projects located at multiple laboratories and universities.

We present some science case studies that indicate that a paradigm shift in the process of science will come about as a result of Grids providing transparent and secure access to advanced and integrated information and technologies infrastructure: powerful computing systems, large-scale data archives, scientific instruments, and collaboration tools. These changes will be in the form of services that can be integrated with the user's work environment, and that enable uniform and highly capable access to these computers, data, and instruments, regardless of the location or exact nature of these resources. These services will integrate transient-use resources like computing systems, scientific instruments, and data caches (e.g., as they are needed to perform a simulation or analyze data from a single experiment); persistent-use resources, such as databases, data catalogues, and archives, and; collaborators, whose involvement will continue for the lifetime of a project or longer.

While we largely address large-scale science in this paper, Grids, particularly when combined with Web Services, will address a broad spectrum of science scenarios, both large and small scale.

1 What is the General Idea of Grids?

Computing, data, and collaboration Grids ([1] [2] [3]) are an approach for building dynamically constructed collaborative problem solving environments using geographically and organizationally dispersed high performance computing and data handling resources.

[a] wejohnston@lbl.gov, www.itg.lbl.gov/~wej, www.ipg.nasa.gov

USMail: Lawrence Berkeley National Laboratory, MS 50B-2239 , Berkeley, CA, 94720, USA

The overall motivation for current large-scale, multi-institutional Grid projects is to enable the resource and human interactions that facilitate large-scale science and engineering such as aerospace systems design [4], high energy physics data analysis [5], climatology [6], large-scale remote instrument operation [7], collaborative astrophysics based on virtual observatories [8], etc. In this context, the goal of Grids is to provide significant new capabilities to scientists and engineers by facilitating routine construction of information and collaboration based problem solving environments that are built on-demand from large pools of resources.

Functionally, Grids will provide tools, middleware, and services for:

- o building the application frameworks that allow discipline scientists to express and manage the simulation, analysis, and data management aspects of overall problem solving
- o providing a uniform look and feel to a wide variety of distributed computing and data resources
- o supporting construction, management, and use of widely distributed application systems
- o facilitating human collaboration through common security services, and resource and data sharing
- o providing remote access to, and operation of, scientific and engineering instrumentation systems
- o managing and securing this computing and data infrastructure as a persistent service

This is accomplished through two aspects: 1) A set of uniform software services that manage and provide access to heterogeneous, distributed resources, and, 2) a widely deployed infrastructure. The software architecture is depicted in Figure 1, and the deployment issues are discussed later.

Discipline Portals / Frameworks (problem expression; user state management; collaboration services; workflow engines; fault management)
Applications and Utility Services (domain specific and general components)
Language Specific APIs (Python, Perl, C, C++, Java)
Grid Collective Services (resource brokering; resource co-allocation; data cataloguing, publishing, subscribing, and location management; collective I/O, job management)
Core Grid Functions (resource discovery; resource access; authentication and security; event publish and subscribe; monitoring / events)
Communication Services
Security Services
Resource Managers (export resource capabilities to the Grid, handle execution environment establishment, hosting, etc., for compute resources)
Physical Resources (computers, data storage systems, scientific instruments, etc.)

Figure 1.   Grid Architecture

2 Application Case Studies

Many large-scale science projects are being forced to deal with various issues such as large distributed data sets, diverse computational resources, and collaboration management. The case studies below highlight the current approach and future requirements of some representative examples of large-scale science projects.

2.1 High Energy and Nuclear Physics: A Data-Intensive Environment [a]

The major high energy physics (HEP) experiments of the next twenty years will break new ground in our understanding of the fundamental interactions, structures and symmetries that govern the nature of matter

and space-time. Among the principal goals are to find the mechanism responsible for mass in the universe, and the "Higgs" particles associated with mass generation, as well as the fundamental mechanism that led to the predominance of matter over antimatter in the observable cosmos.

The largest collaborations today, such as CMS [9] and ATLAS [10] that are building experiments for CERN's Large Hadron Collider program (LHC, [11]), each encompass 2000 physicists from 150 institutions in more than 30 countries. The current generation of operational experiments at Stanford Linear Accelerator Center (SLAC) (BaBar [12]) and FermiLab (D0 [13] and CDF [14]), as well as the experiments at the Relativistic Heavy Ion Collider (RHIC, [15]) program at Brookhaven National Lab, face similar challenges. BaBar, for example, has already accumulated datasets approaching a petabyte [a].

The HEP (or HENP, for high energy and nuclear physics) problems are among the most data-intensive known. Hundreds to thousands of scientist-developers around the world continually develop software to better select candidate physics signals from particle accelerator experiments such as CMS, better calibrate the detector and better reconstruct the quantities of interest (energies and decay vertices of particles such as electrons, photons and muons, as well as jets of particles from quarks and gluons). These are the basic experimental results that are used to compare theory and experiment. The globally distributed ensemble of computing and data facilities (e.g., see Figure 2), while large by any standard, is less than the physicists require to do their work in an unbridled way. There is thus a need, and a drive, to solve the problem of managing global resources in an optimal way in order to maximize the potential of the major experiments to produce breakthrough discoveries.

Figure 2. High Energy Physics Data Analysis

This science application epitomizes the need for collaboratories supported by Grid computing infrastructure in order to enable new directions in scientific research and discovery. The CMS situation depicted here is very similar to Atlas and other HEP experiments. (Adapted from original graphic courtesy Harvey B. Newman, Caltech.)

Collaborations on this global scale would not have been attempted if the physicists could not plan on high capacity networks: to interconnect the physics groups throughout the lifecycle of the experiment, and to make possible the construction of Data Grids capable of providing access, processing and analysis of massive datasets. These datasets will increase in size from petabytes to exabytes (1 EB = 10^{18} bytes) within the next decade. Equally as important is highly capable middleware (the Grid data management
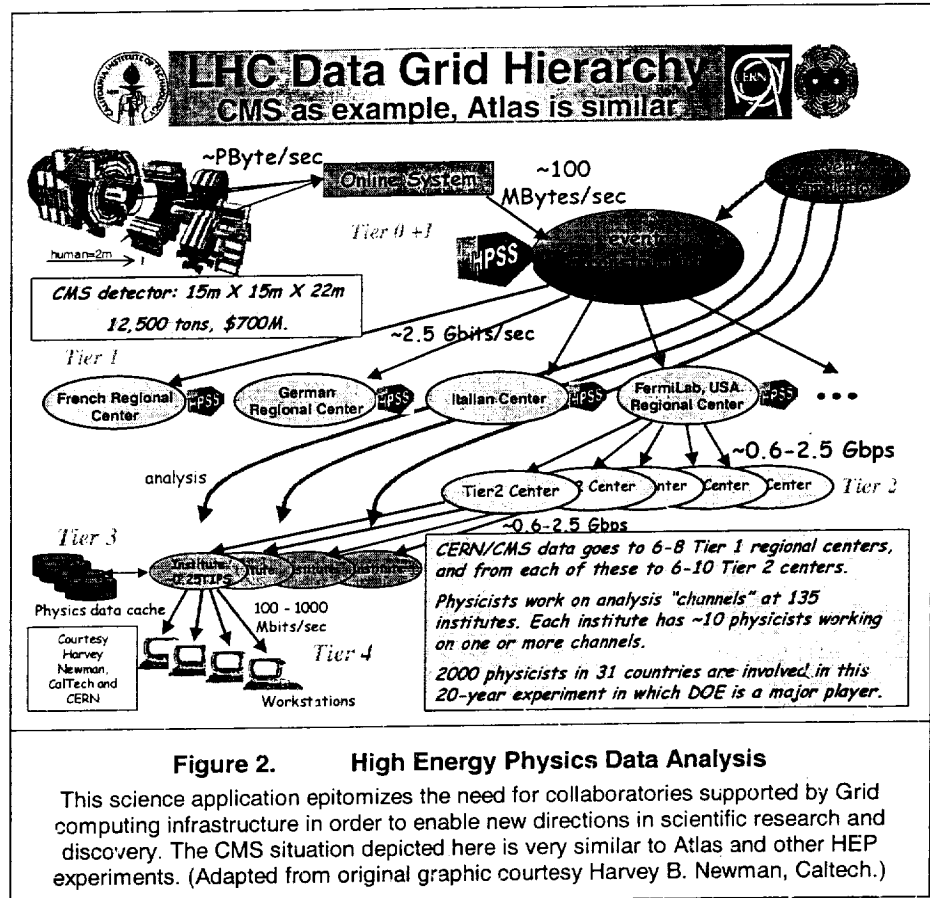
---

[a] 1 petabyte = 1,000 terabytes = 1,000,000 gigabytes = 10^9 megabytes = 10^{15} bytes

and underlying resource access and management services) to facilitate the management of world wide computing and data resources that must all be brought to bear on the data analysis problem of HEP.

Successful construction of network and Grid middleware systems able to serve the global HEP, as well as other scientific communities with data-intensive needs, could have wide-ranging effects: on research, industrial and commercial operations. The key is intelligent, resilient, self-aware, and self-forming systems able to support a large volume of robust terabyte and larger transactions, able to adapt to a changing workload, and capable of matching the use of distributed resources to policies[a]. These systems could provide a strong foundation for managing the large-scale data-intensive operations processes of the largest research organizations, as well as the distributed business processes of multinational corporations in the future.

Several important collaborations are involved in the HEP effort to use Grids for distributed data processing. The DOE Science Grid [18] is working on identifying and resolving the issues for building production Grids for the DOE Office of Science [19]. The Particle Physics Data Grid (PPDG, [20]) -- jointly funded by the DOE/MICS Office [21] and the DOE HENP Office [22] – is working on Grid middleware and systems for distributed analysis of HEP experiment data.

To cite one example of the Grid technology issues being addressed in HEP, we consider the development of virtualized data, coupled with dataset replication management that the commercial sector calls Content Delivery Networks.

The GriPhyN (Grid Physics Network – http://www.griphyn.org) project is a collaboration of computer science and other IT researchers and physicists from the ATLAS, CMS, LIGO [23], and SDSS [24] experiments. The project is focused on the creation of Petascale Virtual Data Grids that meet the data-intensive computational needs of a diverse community of thousands of scientists spread across the globe. The concept of Virtual Data encompasses the definition and delivery to a large community of a (potentially unlimited) virtual space of data products derived from experimental data or from simulations. In this virtual data space, requests may be satisfied via direct access and/or by (re)computation of simulation data on-demand, with local and global resource management, policy, and security constraints determining the strategy used. That is, what is stored in the metadata is not necessarily just descriptions of the data and pointers to that data, but prescriptions for generating the data. Depending on the implementation and service provided by the Virtual Data system, the user may have to take the prescription and explicitly generate that data, or (as is the case in the GriPhyN project) the system itself will generate the data on demand. Once generated, the data will be managed by the replica manager component, and may be cached at one or several locations in the network.

Overcoming this challenge and realizing the Virtual Data concept requires advances in three major areas:

- o Virtual data technologies
   Advances are required in information models and in new methods of cataloging, characterizing, validating, and archiving software components to implement virtual data manipulations / generation.

- o Policy-driven request planning and scheduling of networked data and computational resources
   Mechanisms are required for representing and enforcing both local and global policy constraints and new policy-aware resource discovery techniques.

---

[a] This is in the realm of an emerging field called Recovery Oriented Computing systems [16]. IBM, for example, has a Grid-like project for ROC in distributed computing environments called Autonomic Computing. The Grid Core Functions [17] are intended for provide sufficient functionality and services to support this approach in the distributed environment.

o   Management of transactions and task-execution across national-scale and worldwide virtual organizations

New mechanisms are needed to meet user requirements for performance, reliability, and cost. Agent computing will be important to permit the Grid to balance user requirements and Grid throughput, with fault tolerance.

The GriPhyN project is primarily focused on achieving these fundamental IT advances that are required to create Petascale Virtual Data Grids, but is also working on creating software systems for community use, and applying the technology to enable distributed, collaborative analysis of data. (E.g., see [25].)

These sorts of Data Grid services are fundamental contributions to Grid technology, and they rely on the basic Grid resource management services being deployed and managed as persistent infrastructure. E.g., see [26].

2.2   Climate [a]

To better understand climate change, we need better climate models – and to achieve such models, we need to exhaustively analyze today's models in order to improve them. The cycle of
analysis → improved model → analysis is typical of climate modeling work generally. One thing this is clear is that climate models today are too low in resolution to correctly represent some important features of the climate. It is expected that adequate computing power will be available over the next 5-10 years, but to determine phenomenon like climate extremes (hurricanes [b], drought and precipitation pattern changes [c], heat waves and cold snaps) and other potential changes as a result of climate change [d], better analysis is needed. Currently, analysis is accomplished by transferring the data of interest from the computer modeling site to the climate scientist's institution for various post-simulation analysis tasks. This can be inefficient if the data volume is large, and several strategies to reduce the data volume before transfer have been developed. However, these processes are often ad hoc and need to be improved or rendered moot.

This means that faster networks are needed to access more climate model data more efficiently, together with middleware to facilitate services such as like visualization and collaboration to assist climate scientists in understanding climate models and climate change. Since climate models require large computing resources, there are only a few sites in the U.S. and worldwide that are suitable for executing these models. In addition, for model efficiency reasons, the data produced by these integrations are stored at the same sites – however, climate scientists are scattered all over the world, which means that, like high energy physics, data distribution for analysis is critical.

---

[a] This section is based on material from Gary Strand (strandwg@ucar.edu), National Center for Atmospheric Research, and was adapted from [6].

[b] Hurricane Andrew was almost exactly 10 years ago and cost many lives and about $20 billion damage. Current climate models aren't quite good enough to resolve hurricanes, but research models driven by reasonably realistic future climate scenarios imply that Andrew-strength hurricanes striking the US will become more common. That implies many more billions in damage and more deaths.

[c] Likewise, the drought the Western US is currently facing could become the typical climate pattern, with millions of acres of forests burning in wildfires, and things like the cost of supplying water to the burgeoning populations of the Western US. Changes in precipitation location may also make agriculture in the Midwest US more problematic – either extended dry periods or floods like those that plagued the upper Midwest in the early 1990s.

[d] This refers to changes in disease patterns, for example. It's possible that climate change may make the US more susceptible to the spread of diseases found today mostly in the tropics. The West Nile virus is relatively innocuous compared to malaria.

Over the next five years, climate models will see an even greater increase in complexity than that seen in the last ten years. Influences on climate – input to the models – will no longer be approximated by essentially fixed quantities, but will become simulation components in and of themselves (e.g., see Figure 3). The North American Carbon Project (NACP), which endeavors to fully simulate the carbon cycle, is an example. Increases in resolution, both spatially and temporally, are in the plans for the next two to three years. The atmospheric component of the coupled system will have a horizontal resolution of approximately 150 km and 30 levels. A plan is being finalized for such model simulations that will create about 30 terabytes of data in the next 18 months, which is double the rate of current model data generation, e.g. from the Parallel Climate Model (PCM, [27]).

Figure 3. **There are many complex simulations that interact to produce a comprehensive climate model.**

(Courtesy Gordon Bonan: *Ecological Climatology: Concepts and Applications.* Cambridge University Press, Cambridge, 2002.)
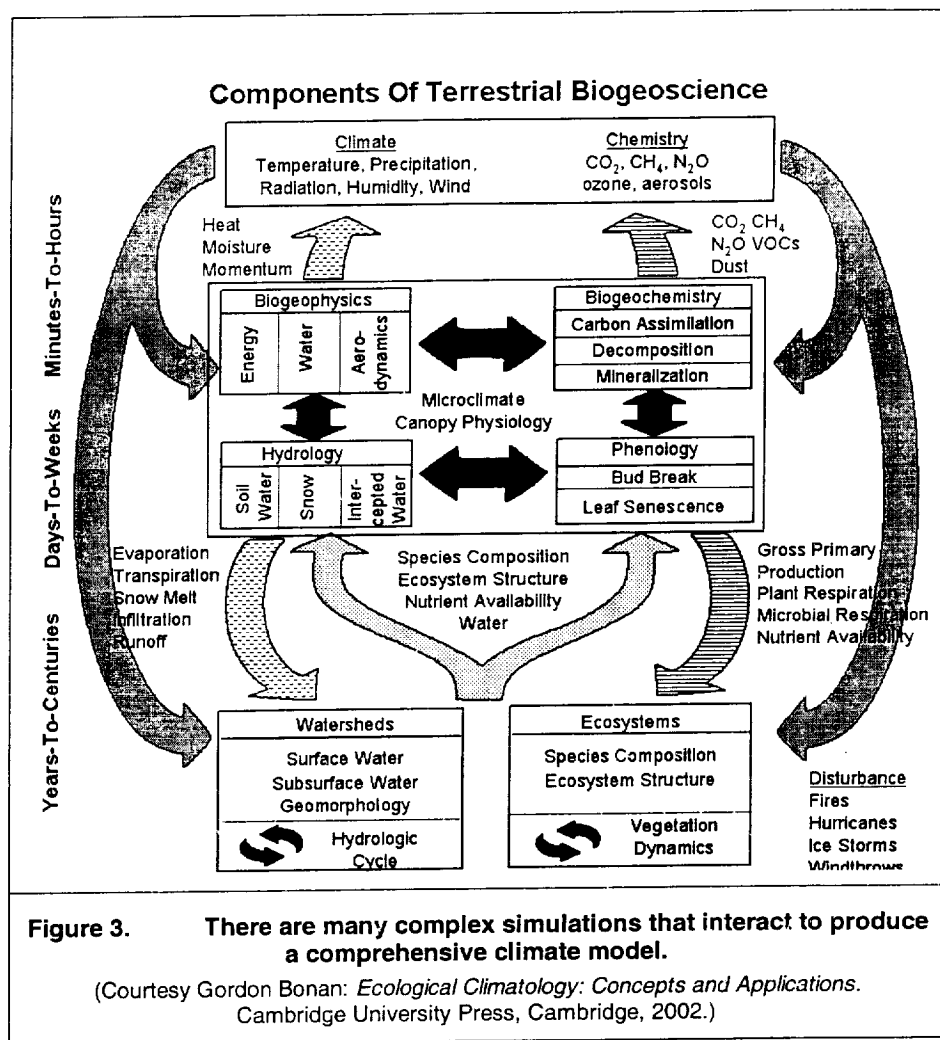
These much finer resolution models, as well as the distributed nature of computing resources, will demand much greater bandwidth and robustness from computer networks than is presently available, and middleware to couple manage and couple the components together. These studies will be driven by the need to determine future climate at both local and regional scales as well as changes in climate extremes - droughts, floods, severe storm events, and other phenomena. Climate models will also incorporate the vastly increased volume of observational data now available (and that will be available in the future), both for hind casting (simulation of past climate) and inter-comparison purposes.

The end result is that instead of tens of terabytes of data per model instantiation, hundreds of terabytes to a few petabytes of data will be stored at multiple computing sites, to be analyzed by climate scientists worldwide. The Earth System Grid [28] and its descendents will be fully utilized to disseminate model data and for scientific analysis. Additionally, these more sophisticated analyses and collaborations will increase the needed network resources and infrastructure. It's expected that considerably more climate scientists will examine the model data than do so today. PCM data has been analyzed by scientists at

UCSD [a], the University of Colorado at Boulder, NOAA [b], NERSC [c], PNNL [d], as well as in Sweden, Germany and Japan. Bulk data transfer will be necessary to support the substantial increases, as well as Grid based remote access tools services.

As climate models become more multidisciplinary, scientists from fields outside of climate, oceanography and the atmospheric sciences will collaborate on the development and examination of climate models. Biologists, hydrologists, economists and others will assist in the creation of additional components that represent important, but as-yet poorly understood, influences on climate. These models, sophisticated themselves, will likely be utilized at computing sites other than where the climate model is executed. In order to maintain efficiency, dataflow to and from these collaboration efforts will demand extremely robust and fast networks and middleware to coordinate the models and techniques to simplify the interconnection of the models.

Beyond five years out, climate models will again increase in resolution, and many more fully simulated components will be integrated. At this time, the atmospheric component may become nearly mesoscale (commonly used for weather forecasting) in resolution, 30 km by 30 km, with 60 vertical levels. . Data volumes could reach several petabytes, which is a conservative estimate. Climate models will be used to drive regional scale climate and weather models, which require resolutions in the tens to hundreds of meters range, instead of the typical hundreds of kilometers resolution of the CCSM [e] and PCM. There will be a true carbon cycle component, models of biological processes will be used, for example, simulations of marine biochemistry (which affects the interchange of greenhouse gases like methane and carbon dioxide with the atmosphere), and fully dynamic vegetation. These scenarios will include human population change and growth (which effect land usage and rainfall patterns) and econometric models, to simulate the potential changes in natural resource usage and efficiency. Additionally, models representing solar processes, to better simulate the incoming solar radiation, will be integrated. Climate models at this level of sophistication will likely be run at more than one computing center in distributed fashion, which will demand extremely high speed and very robust computer networks to interconnect them. together with very sophisticated middleware to facilitate the integration of all of these models which are likely to be running at the sites where the expertise resides. This circumstance is common, e.g., in the aerospace design community: models and associated engineering databases are maintained by a small group of specialists at their home institutions, When the model and data are needed, they are provided as a remote service (increasingly a Grid service). This is where the Grid middleware provides the necessary access and integration services. The coupling and integration of models will be facilitated by the new integration of Web Services and Grids, described below.

2.3 Magnetic Fusion Energy [f]

The long-term goal of magnetic fusion research is to develop a reliable energy source that operates on the same general principles as those of the Sun, and that is environmentally and economically sustainable. To achieve this goal, it is necessary to develop the science of plasma physics, a field with close links to fluid mechanics, electromagnetism, and non-equilibrium statistical mechanics. The highly collaborative nature

---

[a] San Diego Supercomputer Center [29]

[b] U.S. National Oceanic and Atmospheric Administration

[c] U. S. Dept. of Energy, Office of Science, National Energy Research Scientific Computing Center [30] located at Lawrence Berkeley National Laboratory [31]

[d] Pacific Northwest National Laboratory [32]

[e] Community Climate System Model [33]

[f] This section is based on material from D.P. Schissel, General Atomics Fusion Group (schissel@fusion.gat.com), M.J. Greenwald, MIT Plasma Science and Fusion Center (G@PSFC.MIT.EDU), and W.E. Johnston, Lawrence Berkeley National Laboratory, and was adapted from [6].

of the Magnetic Fusion Energy Sciences (MFE) is due to the small number of experimental facilities (see Figure 4) and a computationally intensive theoretical program, are creating new and unique challenges for computer networking and middleware.

In the United States, experimental magnetic fusion research is centered at three large facilities (Alcator C–Mod [34], DIII–D [35], NSTX [36]) with a present day replacement value of over $1B; clearly too expensive to duplicate. As these experiments have increased in size and complexity, there has been concurrent growth in the number and importance of collaborations between large groups at the experimental sites and associated groups located at universities, industry sites, and national laboratories.

Teaming with the experimental community is a theoretical and simulation community whose efforts range from the very applied analysis of experimental data, too much more fundamental theory like the creation of realistic non–linear 3D plasma models. The MFE simulation community is one of the largest users of scientific supercomputing resources in the U.S.

The three main magnetic fusion experimental sites operate in a similar manner. The gross tokamak machine hardware parameters are configured before the start of the experimental day. Magnetic fusion experiments operate in a pulsed mode producing plasmas of up to 10 seconds duration every 10 to 20 minutes, with 25–35 pulses per day. For each plasma pulse up to 10,000 separate measurements versus time are acquired at sample rates from kHz to MHz, representing hundreds of megabytes of data.

Throughout the experiment session, hardware/software plasma control adjustments are made as required by the experimental science. These adjustments are debated and discussed amongst the experimental team (typically 20–40 people) with most working on site in the control room but with many participating from remote locations. Decisions for changes to the next plasma pulse are informed by data analysis conducted within the roughly 15 minute between-pulse interval. This mode of operation places a large premium on rapid data analysis that can be assimilated in near–real–time by a geographically dispersed research team.

The computational emphasis in the experimental science area is to perform more and more complex data analysis between plasma pulses. For example, today a complete time–history of the plasma magnetic structure is available between pulses by using parallel processing on Linux clusters. Five years ago, only selected times were analyzed between pulses with the entire time–history completed overnight. Five years from now, analysis that is today performed overnight should be completed between pulses. Such enhanced between-pulse data analysis will include more advanced simulations that will run on large-scale computing resources that are remote from the experiment. The ability to more accurately compare

experiment and theory between pulses will greatly enhance the value of experimental operations. Today, these comparisons are done after experimental operations have concluded when it is too late to adjust experimental conditions. This is very limiting for the experimentalists who typically only get a few days a year on a fusion device to test out their theories.

With the creation of more data between pulses there exists an increasing burden to assimilate all of the data. Enhanced visualization tools are presently being developed that will allow this order of magnitude increase to be effectively used for decision making by the experimental team. Clearly, the movement of this quantity of data in a 15–20 minute time window to computational clusters, to data servers, and to visualization tools used by an experimental team distributed across the United States and other countries, and with ITER, around the world. Clearly, the sharing of remote visualizations back into the control room
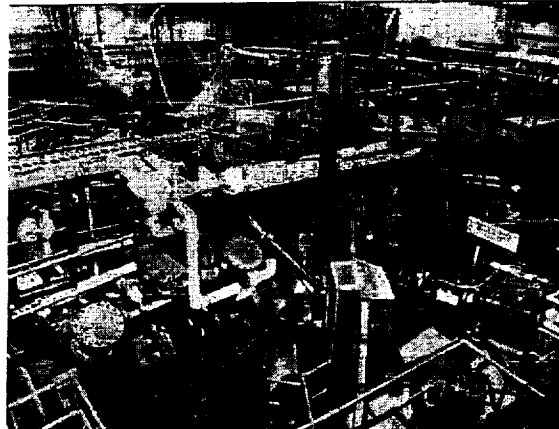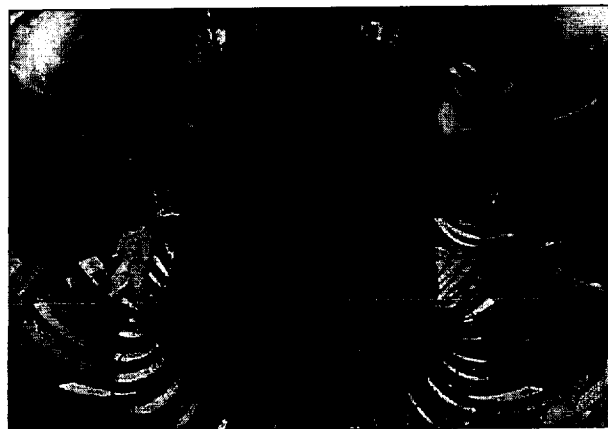
Figure 4.   Tokamak Magnetic Fusion reactors are large, complex, and expensive. There are only a few in the world for fusion energy experiments.

Top left: Human inside a tokamak. Top right: The environment of the DIII-D tokamak at General Atomics, San Diego, CA, (note the human on the catwalk on the left side). (From "Creating a Star on Earth" http://fusioned.gat.com/Teachers/Teachers.html.) Bottom right: Drawing of the planned ITER - International Thermonuclear Experimental Reactor (note human figure at bottom for scale).
From http://www.iter.org/
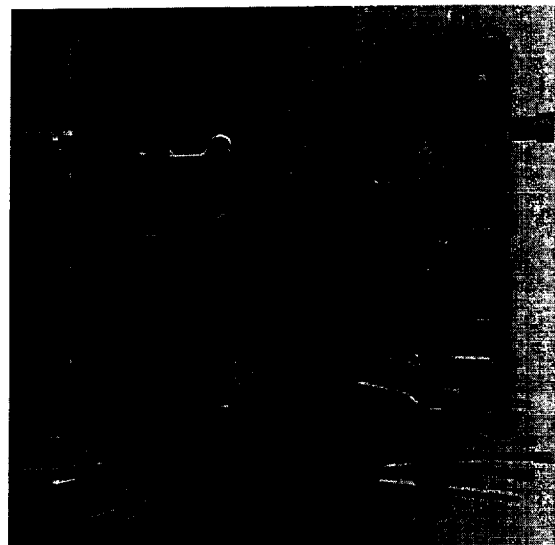
will place a severe burden on present day network and middleware technology.

Although the fundamental laws that determine the behavior of fusion plasmas are well known, obtaining their solution under realistic conditions is a computational science problem of enormous complexity.

Datasets generated by these simulation codes will approach the 1 TB level within the next three to five years. Additionally, these datasets will be analyzed like experimental plasmas are analyzed to extract further information. Therefore, the data repository for simulations will be dynamically evolving rather than a write–once type scenario. These large datasets will most likely be dispersed across the collaborator sites and will be made available using various data Grid-like services.

In addition to the network bandwidth requirements implied above, the nature of MFE research also leads to requirements for advanced middleware services. As in other sciences, valuable resources such as computers, data, instruments and people are distributed geographically and must be shared for successful collaboration. In fusion, the need for real-time interactions among large experimental teams and the requirement for interactive visualization and processing of very large simulation data sets are particularly challenging.

In terms of Grid services, for example, the apparently conflicting requirements for transparency and security in a widely distributed environment point up the need for efficient and effective services in this area. Central management of authentications (PKI or equivalent technologies) using "best practices" and providing 24×7 support is essential. Further, it is essential that the user authentication framework and operational environments are such that common policy may be negotiated among international collaborators in order to enable collaborations to span international boundaries and between application development and site security groups. Development of mutually agreed upon tools and protocols for resource authorization is equally important.

As fusion collaboratory activities grow, the needs for global data and collaboration directory and naming services will expand as well. A hierarchical infrastructure with well–managed "roots" can provide the necessary glue for many collaborative activities. Analogous to the Internet's domain name services, this infrastructure would give local resource managers needed flexibility while maintaining global connectivity and persistence. A global name service could even solve the longstanding problem in the field of computational simulation variable name translation between codes or experiments. Grid services for queuing and monitoring in the distributed computing environment are also needed. These must be easy to configure and deploy and robust in operation.

2.4   Data-Driven Astronomy and Astrophysics [a]

Technological advances in telescope and astronomy instrument design during the last ten years, coupled with the exponential increase in computer and communications capability, have caused a dramatic and irreversible change in the character of astronomical research.

Formerly, individual astronomers requested observing time on an instrument in order to study a few specific objects or a small region of the sky. Today, the instruments are so big and expensive that this is not practical. This has lead to a paradigm shift in how astronomy is being done, and at the same time it has vastly expanded the potential for new and discovery-based astronomy.

Many new instruments are essentially being run all the time, taking as many observations as possible, over as much of the sky as possible. Large-scale surveys of the sky from space and ground are being initiated at wavelengths from radio to X-ray, thereby generating vast amounts of high-quality data. These surveys are creating catalogs of objects (stars, galaxies, quasars, etc.) numbering in billions, with up to a hundred measured parameters for each object. Yet this is just a foretaste of the much larger data sets to come. Astronomy is being done on the collected data sets rather than through direct use of the instrument.

---

[a] This section is based on material from the Virtual Observatories of the Future conference (http://www.astro.caltech.edu/nvoconf/), from the National Virtual Observatory white paper, also at that location, and from contributions by Julian Borrill (LBNL/NERSC, JDBorrill@lbl.gov) and Paul Messina, (CalTech, messina@cacr.caltech.edu).

Further, this mode of operation allows for an unprecedented simultaneous analysis of high-quality observations from many instruments with different characteristics observing the same part of the sky. This has already led to some important science results that would not have been possible with single instrument observation.

This new paradigm will enable tackling some major astronomy problems with an unprecedented accuracy. High-quality coverage over large parts of the sky in multiple wavelengths will provide data on billions of objects, and will allow discovery of new phenomena (from the analysis of statistically rich and unbiased image databases) and understanding of complex astrophysical systems (through the interplay of data and simulation). It will permit the discovery of rare objects (e.g., at the level of one source in 10 million) that may well lead to surprising new discoveries of previously unknown types of objects or new astrophysical phenomena, and it will permit the multi-wavelength identification of large statistical samples of

An example of the different signal sources that must be taken into account in an observation of the Cosmic Microwave Background: detector noise, dust, synchrotron, free-free, galaxies, kinetic Sunyaev-Zel'dovich, thermal Sunyaev-Zel'dovich, and the CMB itself. Understanding the impact of each of these on the total observation requires high quality data at a range of frequencies from 10GHz to 1000GHz. (Image courtesy Julian Borrill, LBNL/NERSC)

Figure 5. **The cosmic microwave background power spectrum supports the model of a flat Universe.**

previously rare objects (brown dwarfs, high-z quasars, ultra-luminous IR galaxies, etc.) For example, see "New Science: Rare Object Searches" in [37]. This large coverage, periodically repeated, will allow cross-identification of "unidentified sources" (e.g., using radio, optical, and IR surveys to identify serendipitous Chandra X-ray sources), and it will allow identification of targets for specific spectrographic follow-up, as is done in supernova cosmology. The data will also provide for mapping of the large-scale structure of the universe.

Periodic re-surveys will allow for the discovery of objects and phenomena that change on observational time scales. Given that human observational time scales are minuscule on a cosmic scale, these events tend to represent something fairly dramatic. Examples include near-Earth asteroids, supernovae, gamma ray bursts, pulsars, etc.

Another class of query uniquely enabled by the multi-instrument sky surveys, and of direct relevance to understanding the fundamental structure of matter, will be searches for information at all wavelengths on a particular region of the sky. As astronomers attempt to detect fainter and fainter signals, such searches will become increasingly important. For example, the spectrum of anisotropies in the polarization of the cosmic microwave background radiation (Figure 5) is sensitive to gravitational wave emission during the inflation of the early universe, and hence probes physics at the Grand Unified Theory energy scale—energies beyond the capability of any imaginable accelerator. However, this signal is extremely faint and
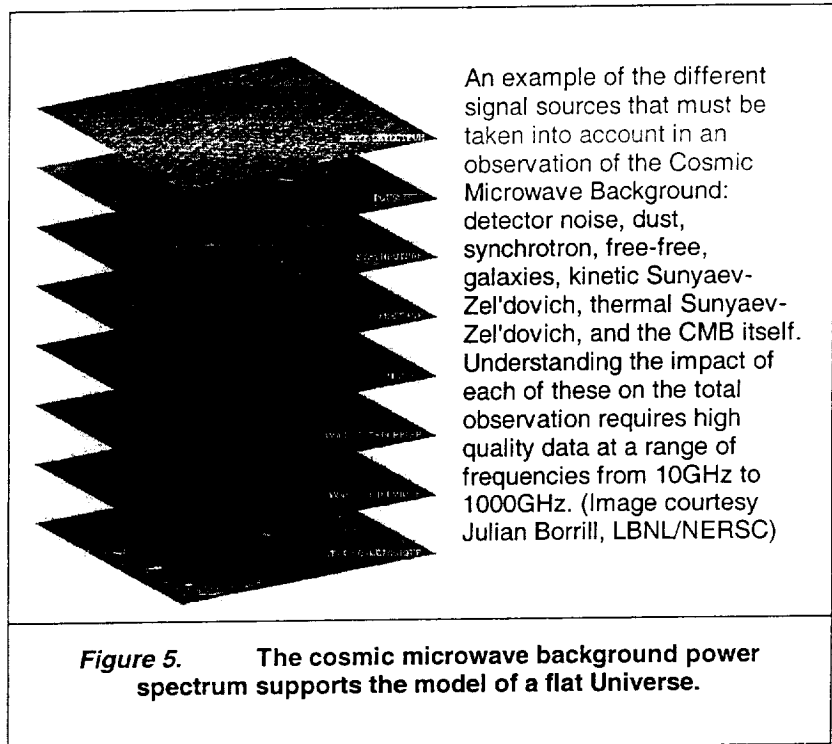
as yet undetected[a]. Obtaining such a measurement will require detailed understanding of all possible foreground sources (see Figure 5).

This sort of "virtual" astronomy involves accessing 20-40 major astronomical databases around the world, and the required joint searches of the surveys encompassed in projects like the National Virtual Observatory (NVO) [39] is critical for astronomers, both to select regions of the sky with as little contamination as possible in advance of an observation, and to characterize the location and spectral dependency of whatever sources there were afterwards. These searches require extracting large amounts of data, and then moving the multiple datasets to computational facilities for the required extraction and multi-instrument rectification needed for cross dataset (cross observation) comparisons. The NVO is using Grid technology to access and analyze these very large, distributed datasets. (See Figure 6 and the project description [38].)

Figure 6.   The NVO Architecture

"The correspondence of the NVO architecture layers to the Grid infrastructure layers is shown on the right side of the diagram. Each component is designed to support access to the existing survey digital libraries and to the expanded capabilities required by the NVO to support analyses that require processing of a large fraction of the catalog holdings or images from multiple surveys." (From the NVO Project Description [38])

These types of scientific investigations were not feasible with the more limited datasets of the past: We are at the start of a new era of information-rich astronomy. Large digital sky surveys and data archives are becoming the principal sources of data in astronomy. The very style of observational astronomy is changing: systematic sky surveys are now used both to answer some well-defined questions which require large samples of objects, and to discover and select interesting targets for follow-up studies with space-based or large ground-based telescopes. However, this vision relies completely on well-developed and highly capable software, computing, and networking infrastructure, and Grid software that is being deployed to address the middleware issues.

---

[a] In a press release dated 19 Sept. 2002, John E. Carlstrom (U. Chicago) and his team announced the they had observed the polarization of the Cosmic Microwave Background using the Degree Angular Scale Interferometer (DASI) instrument operating at the South Pole. See http://astro.uchicago.edu/dasi/.

3 Advanced Infrastructure as an Enabler for Future Science

The science case studies in the previous section give an indication of the future process of science that would require, or is enabled by, significant increases in computing and networking capacity, and middleware functionality.

Several general observations and conclusions may be made after analyzing these application scenarios.

The first, and perhaps most significant, observation is that a lot of science is already, or is rapidly becoming, an inherently distributed endeavor: Science experiments involve a collection of collaborators that are frequently multi-institutional, the data and computing requirements are routinely addressed with compute and data resources that are frequently even more widely distributed than the collaborators, and as scientific instruments become more and more complex (and therefore more expensive) they are frequently used as shared facilities with remote users. Even numerical simulation – an endeavor previously typically centered on one, or a few, supercomputers – is becoming a distributed endeavor. Simulations are increasingly producing data of sufficient fidelity that it is used in post-simulation situations: As input to other simulations, to guide laboratory experiments, or to compare with other approaches to the same problem to motivate competitive improvements of the underlying models. This sort of science depends critically, or will in the near future, on an infrastructure that supports the process of distributed science.

A second observation is that when asked what sort of services are needed to support distributed science, the answer always involves a lot of middleware services beyond just basic computing and networking capacity.

A third observation is that there is considerable commonality in the services needed by the various science disciplines. This means that we can define a common "infrastructure" for distributed science

Fourth, all of the science areas need high-speed networks and advanced middleware to couple, manage, and access the widely distributed, high-performance computing systems, the many medium-scale systems of the scientific collaborations, high data-rate instruments, and the massive data archives that, together, are critical to next generation science, and to support highly interactive, large-scale collaboration. All of these elements operating smoothly together are required in order to produce an advanced distributed computing, data, and collaboration infrastructure for science that will enable paradigm shifts in how science is conducted. That is, paradigm
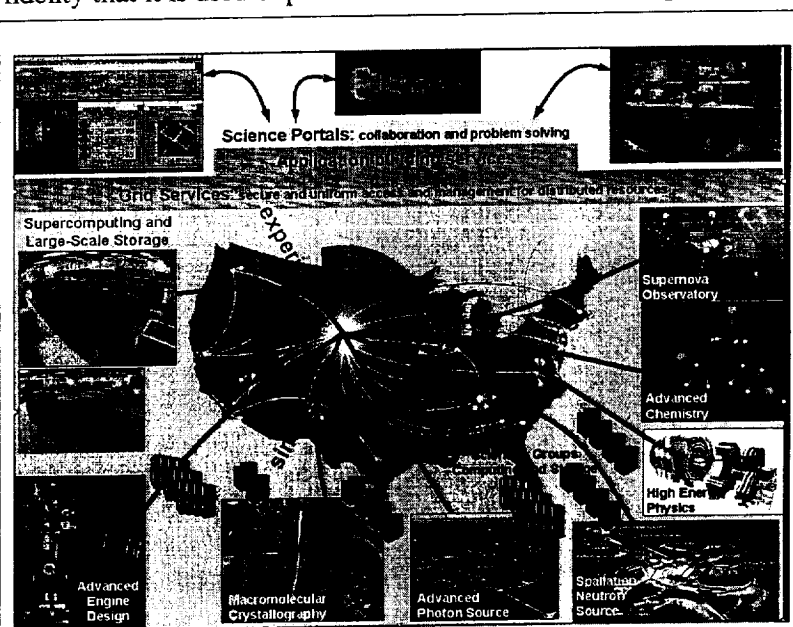
Figure 7.   Integrated Cyber-Infrastructure Enables Advanced Science:
A Vision for the U. S. Dept. of Energy, Office of Science
o Provide the science community with advanced distributed computing infrastructure based on large-scale computing, high speed networking, and Grid middleware
o Enable the collaborative and interactive use of the next generation of massive data producing scientific instruments
o Facilitate large-scale scientific collaborations that integrate the Federal Labs and Universities

shifts resulting from increasing the scale and productivity of science depend completely on such an *integrated advanced infrastructure* that is substantially beyond what we have today. Further, these paradigm shifts are not speculative. Several areas of science are already pushing the existing infrastructure to its limits in trying to move to the next generation of science.

There is a clear trend toward the need for services that allow distributed science activities to scale up in several ways: in the number of participants in a distributed collaboration, the amount of data that can be managed, the diversity of the use of data, the number of people who can discover and use the data, the number of independent computational simulations that can be combined in order to represent more realistic or complex phenomenon or physical system, etc.

The task of the integrated advanced infrastructure is to deliver an overall computing, data, and collaboration quality of service to scientific projects. That is:
- Computing capacity adequate for a task is provided at the time the task is needed by the science,
- Data capacity sufficient for the science task is provided independent of location, and in a transparently managed, global name space,
- Communication capacity sufficient to support all of the aforementioned is provided transparently to both systems and users, and
- Software services supporting a rich environment that lets scientists focus on the science simulation and analysis aspects of software and problem solving systems, rather than on the details of managing the underlying computing, data, and communication resources.

All of these are (or will be) provided by Grid middleware as the mechanism for coupling computing, data, instruments, and human collaborators into an integrated science environment.

3.1 Grid Middleware

The evolution of middleware and distributed systems in the scientific computing environment is currently embodied in computing and data Grids .

As noted above, Grid middleware provides services for uniform access, management, control, monitoring, communication, and security to application developers using these distributed resources. Grid managed resources are the geographically distributed, architecturally and administratively heterogeneous computing, data, and instrument systems of the scientific milieu. That is, the role of Grid middleware is to greatly simplify the construction and use of widely distributed and/or large-scale collaborative problem solving systems that are using these resources.

The international group working on defining and standardizing Grid middleware is the Global Grid Forum ("GGF," [40]) that now consists of some 700 people from some 130 academic, scientific, and commercial organizations in about 30 countries. GGF involves both scientific and commercial computing interests. It also entails an evolving understanding of the issues that must be addressed in order to facilitate the expeditious construction of the complex distributed systems of science from a very dynamic pool of resources.

There is now enough experience in building Grids (e.g. DOE Science Grid, NASA's IPG [41], the UK eScience Grid [42], EU DataGrid [43], etc.) that the basic access and management functions noted above are fairly well understood, and reference implementations are available for most of these through the Globus toolkit [44].

However, as our experience with Grids grows more issues arise that must be addressed in order to meet the goals of easily building effective distributed science systems.

In order to be effective, interoperable Grid middleware must be widely deployed. This involves two things. First, it must be recognized that Grids represent an essential new aspect of the infrastructure of science, and thus must be supported as persistent infrastructure. The issues of operating Grids as production infrastructure as discussed in [45] and [26]. Second, an educational process must address the critical sociological issues involved in modifying operational procedures, inter-site cooperation and sharing, homogenizing security policy etc., as the institutional groups that deal with these issues start to embrace Grids. Many of these issues have been addressed in the narrower scope of building and operating networks, and now have to be addressed in the broader scope of interoperating of computing, data, and instrumentation facilities.

The type of Grid middleware described thus far provides the essential and basic functions for resource access and management. As we deploy these services and gain experience with them, it has also become clear that higher level services are required in order to make effective use of distributed resources. These higher-level services include, e.g., functionality such as brokering to automate building application-specific virtual systems from large pools of resources and collective scheduling of resources so that they may operate in a coordinated fashion. (That is, so that a high performance computing system could do the real-time data analysis that would enable a scientist to interact with experiments involving on-line instruments or to allow simulations from several different disciplines to exchange data and cooperate to do a whole system simulation, as is increasingly needed to study real, complex physical and biological systems.) These types of services are currently being developed and/or designed.

Higher level services also provide functionality that aids in componentizing and composing different software functions so that complex software systems may be built in a "plug-and-play" fashion. These services are being approached by leveraging large industry efforts in XML based Web Services^a by integrating Web Services and Grid services. This will allow the use of commercial and public domain tools such as Web interface builders, problem solving environment framework builders, etc., to build the complex application systems that provide the rich functionality needed for maximizing human productivity in the practice of science. It will also provide for describing the interfaces and data of scientific simulations, and while the interfaces and data types of science tend to be more complex than those of commerce (e.g. XML primitive data types only represent a subset of the data types of science), this should still prove useful in addressing some aspects of the problem of coupling simulations. This Web-Grid integration (see [48], [49], [50]) is currently a major thrust at the Global Grid Forum in the form of the Open Grid Services Interface Working Group [51], and while much work remains , the potential payoff for science is considerable. (E.g., see [52] and [53].)

3.2 Platform Services

Another aspect of the middleware is the support that is needed on the resource platforms themselves.

Computing system must have schedulers that enable co-scheduling with other, independent resources. Data archive systems must have access servers that allow for reliable, high-speed, wide-area network data transfer. Networks must provide capabilities for quality-of-service (usually in the form of bandwidth guarantees) that let distributed resources communicate at high bandwidth during critical times in coupled simulation or on-line instrument data analysis. All of the storage, computing, and network resources must have support for the detailed monitoring that is essential for debugging and fault detection and recovery in widely distributed systems.

---

^a Web services are a set of industry standards being developed and pushed by the major IT industry players (IBM, Microsoft, Sun, Compact, etc.). They provide a standard way to describe and discover Web accessible application components, and a standard way to connect and interoperate these components. See, e.g., [46], [47],

These services must be developed, installed, and integrated into the operational environments of all of the individual systems that make up the resource pools of science.

3.3 Grid Middleware Conclusions

Grid middleware has thus far shown considerable promise toward providing the resource integration required by distributed science. (E.g., see Figure 8.) However, we are just at the beginning of the development and deployment of Grid middleware, and Grids are actively in the process of evolution.

Grids are currently focused on resource access and management. This is a necessary first step to provide a uniform underpinning, but is not sufficient if we are to realize the potential of Grids for facilitating science and engineering. Unless an application already has a framework that hides the use of these low level services (which was the case in several of the examples above), the Grid is difficult to use for most users. To address this, Grids are evolving to a service oriented architecture.

Users are primarily interested in "services" – software modules that perform functions directly useful to their science, such as a particular type of simulation, or a broker that finds the "best" system to run a job. Even many Grid tool developers, such as those that develop application portals, are primarily interested in services – resource brokering, workflow management, user security credential management, etc. This is an area where much more work is needed.

The IT industry expects that most, if not all, of it's applications to be packaged as Web services in the future, and the evolution of Grids toward services is going hand-in-hand with a large IT industry push to develop an integrated framework for Web services.

The integration of Grids with Web services also addresses several missing capabilities in the current Web Services approach (e.g. creating and managing task instances). It will also provide for more easily integrating commercial software/services with scientific and engineering applications and infrastructure.
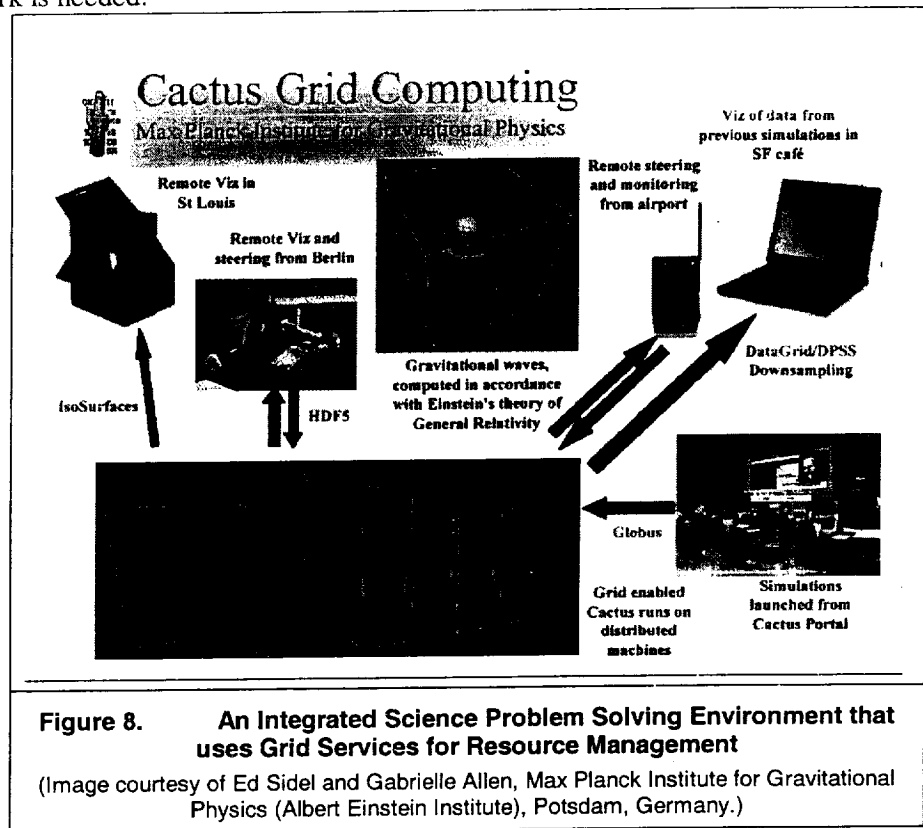
Figure 8. **An Integrated Science Problem Solving Environment that uses Grid Services for Resource Management**

(Image courtesy of Ed Sidel and Gabrielle Allen, Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Potsdam, Germany.)

In summary, the goal of Grids is to provide significant new capabilities to scientists and engineers by facilitating routine construction of large-scale information-based and collaboration-based problem solving environments that are built on-demand from large pools of shared resources.

Finally, returning to the initial view of Grids, we now recast the architecture (Figure 1) in terms of a set of capabilities that are needed by the various Grid users: The discipline scientists, the science framework builders, the computational scientists, the Grid developers, and the Grid resource managers. These capabilities are indicated in Figure 9.

Capabilities to support scientists / engineers / domain problem solvers

–Collaboration tools (work group management, document sharing and distributed authoring, sharing application session, human communication)
–Programmable portals - facilities to express, manipulate, preserve the representation of scientific problem solving steps (e.g., AVS, MatLab, Excel, SciRun)

–Data discovery (super SQL for globally distributed data repositories), management, mining, cataloguing, publish
–Human interfaces (PDAs, Web clients, high-end graphics workstations)
–Tools to build/manage dynamic virtual organizations

Capabilities to support building the portals / frameworks / problem solving environments

–Resource discovery, brokering, job management
–Workflow management
–Grid management – fault detection and correction
–Grid monitoring and information distribution – event publish and subscribe

–Grid security
–Security and authorization,
–CORBA, Legion-G, Web Services (service discovery, composition, reusable components)

Capabilities to support building and using computational models

–Utilities for visualization, data management (global naming, location transparency (replication mgmt, caching), metadata management, data curation, discovery mechanisms)
–Support for programming on the Grid (Grid MPI, Globus I/O, Grid debuggers, programming environments, e.g.,to support the model coupling frameworks, and to), Grid program execution environment

–User services (documentation, training, evangelism)

Capabilities to support building Grid systems

–Quality of Service functions
–Authorization and allocation management,and accounting systems

–Basic data access and transport
–Grid information service

Communications

Distributed Resources

Operating the Grid / Resource management

–Grid enabled resources – basic access (start processes, data servers) advanced scheduling support, support for monitoring, security and authentication, systems management access
–Resource-level security support for grid security access policy

–Identity management (run the CAs)
–Grid trouble tickets, system status monitoring, configuration management
–Dynamic registry

Figure 9.   **Capabilities for Various Grid Users**

5 Acknowledgements

6 References

[1]   Computational and Data Grids in Large-Scale Science and Engineering, W. Johnston. Future Generation Computer Systems, 2002. http://www.itg.lbl.gov/~johnston/Grids/homepage.html#JFCS2002

[2]   The Anatomy of the Grid: Enabling Scalable Virtual Organizations. I. Foster, C. Kesselman and S. Tuecke. International J. Supercomputer Applications, 2001. 15(3). http://www.globus.org/research/papers.html#anatomy

[3]   *The Grid: Blueprint for a New Computing Infrastructure*, I. Foster and C. Kesselman, eds. 1998, Morgan Kaufmann. http://www.mkp.com/books_catalog/1-55860-475-8.asp

[4]   NASA's Information Power Grid, W. Johnston. http://www.itg.lbl.gov/~johnston/Grids/homepage.html#IPGIT

[5]   GriPhyN (Grid Physics Network). http://www.griphyn.org

[6]   High Performance Network Planning Workshop, DOE Office of Science. August 13-15, 2002: Reston, Virginia, USA. http://doecollaboratory.pnl.gov/meetings/hpnpw

[7]   Network for Earthquake Engineering Simulation Grid (NEESgrid). http://www.neesgrid.org/

[8]   Virtual Observatories of the Future, Caltech. http://www.astro.caltech.edu/nvoconf/

[9]   The Compact Muon Solenoid Technical Proposal, CMS. http://cmsdoc.cern.ch/

[10]   The ATLAS Technical Proposal, Atlas. http://atlasinfo.cern.ch/ATLAS/TP/NEW/HTML/tp9new/tp9.html

[11]   The Large Hadron Collider Project, LHC. http://lhc.web.cern.ch/lhc/general/gen_info.htm

[12]   The BaBar Experiment at SLAC, BaBar. http://www-public.slac.stanford.edu/babar/

[13]   The D0 Experiment at Fermilab, D0. http://www-d0.fnal.gov/

[14]   The CDF Experiment at Fermilab, CDF. http://www-cdf.fnal.gov/

[15]   The Relativistic Heavy Ion Collider at BNL, RHIC. http://www.bnl.gov/RHIC/

[16]   Recovery Oriented Computing (ROC): Motivation, Definition, Techniques, and Case Studies, D. Patterson, A. Brown, P. Broadwell, G. Candea, M. Chen, J. Cutler, P. Enriquez, A. Fox, E. Kiciman, M. Merzbacher, D. Oppenheimer, N. Sastry, W. Tetzlaff, J. Traupman and N. Treuhaft. March 15, 2002, Computer Science Division, University of California at Berkeley. http://roc.cs.berkeley.edu/papers/ROC_TR02-1175.doc

[17]   Core Functions for Production Grids, G. P. A. W. G. Global Grid Forum. http://www-itg.lbl.gov/GPA/#CoreGridFunc

[18]   DOE Science Grid. http://www.doesciencegrid.org

[19]   Office of Science, U.S. Dept. of Energy. http://www.er.doe.gov/

[20]   Particle Physics Data Grid, PPDG. http://www.ppdg.net/

[21]   Mathematical, Information, and Computational Sciences (MICS), Advanced Scientific Computing Research, Office of Science, U.S. Dept. of Energy, MICS. http://www.sc.doe.gov/ascr/mics/

[22]   Office of High Energy and Nuclear Physics (HENP), Office of Science, U.S. Dept. of Energy, HENP. http://www.er.doe.gov/henp/

[23]   Laser Interferometer Gravitational Wave Observatory, LIGO. http://www.ligo.caltech.edu/

[24]   Sloan Digital Sky Survey, SDSS. http://www.sdss.org/

[25]   GriPhyN Virtual Data Toolkit, VDT. http://www.lsc-group.phys.uwm.edu/vdt/

[26]   Implementing Production Grids, W. E. Johnston, in *Grid Computing: Making the Global Infrastructure a Reality*, F. Berman, G. Fox and T. Hey, Editors. 2003, John Wiley. http://www.itg.lbl.gov/~johnston/Grids/homepage.html#Implement

[27]   Parallel Climate Model, PCM. http://www.cgd.ucar.edu/pcm/

[28]   Earth System Grid, ESG. http://www.earthsystemgrid.org/

[29]   San Diego Supercomputer Center, SDSC. http://www.sdsc.edu/

[30]   National Energy Research Scientific Computing Center, NERSC. www.nersc.gov

[31]   Lawrence Berkelry National Laboratory, U. S. Dept. of Energy, LBNL. http://www.lbl.gov/

[32]   Pacific Northwest National Laboratory, U. S. Dept. of Energy, PNNL. http://www.pnl.gov/

[33]   Community Climate System Model, CSSM. http://www.ccsm.ucar.edu/

[34]   The Alcator C-Mod Tokamak Fusion Research Project, Alcator C-Mod. http://www.psfc.mit.edu/cmod/

[35]   DIII–D National Fusion Facility at General Atomics, DIII-D. http://web.gat.com/diii-d/

[36]   The National Spherical Torus Experiment, NSTX. http://nstx.pppl.gov/

[37]   Astro-IT challenges and big UK surveys, A. Lawrence. In *Virtual Observatories of the Future*. 2000. Caltech.http://www.roe.ac.uk/wfau/nvo/index.htm

[38]   Building the Framework for the National Virtual Observatory, P. Messina and A. Szalay. http://www.us-vo.org/docs/nvo-proj.doc

[39]   National Virtual Observatory, NVO. http://www.us-vo.org/

[40]   Global Grid Forum, GGF. http://www.gridforum.org/

[41]   NASA's Information Power Grid, IPG. http://www.ipg.nasa.gov

[42]   UK eScience Program. http://www.research-councils.ac.uk/escience/

[43]   European Union DataGrid Project, E. DataGrid. www.eu-datagrid.org/

[44]   The Globus Project, Globus. http://www.globus.org

[45]   Implementing Production Grids, W. Johnston. http://www.itg.lbl.gov/~johnston/Grids/homepage.html#Implement

[46]   *Building Web Services with Java: Making Sense of XML, SOAP, WSDL, and UDDI*, S. Graham, S. Simeonov, T. Boubez, D. Davis, G. Daniels, Y. Nakamura and R. Neyama. 2002: Sams Publishing Co.

[47]   Web Services, IBM. http://www-106.ibm.com/developerworks/webservices/

[48]   The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration, I. Foster, C. Kesselman, J. Nick and S. Tuecke. http://www.globus.org/research/papers.html#OGSA

[49]   Project Spitfite - Towards Grid Web Service Databases, W. H. Bell, D. Bosio, W. Hoschek, P. Kunszt, G. McCance and M. Silander. 2002. http://www.cs.man.ac.uk/grid-db/papers/ggf5-spitfire.pdf

[50]   Web Services Data Grid Architecture, C. Watson. Jefferson Lab. http://www.jlab.org/hpc/datagrid/WebServicesDataGridArch-v2.pdf

[51]   Open Grid Service Interface Working Group, Global Grid Forum. http://www.gridforum.org/ogsi-wg/

[52]   XCAT 2.0: A Component-Based Programming Model for Grid Web Services, M. Govindaraju, S. Krishnan, K. Chiu, A. Slominski, D. Gannon and R. Bramley. June, 2002, Indiana University, Bloomington, Indiana. http://www.extreme.indiana.edu/xcat/

[53]   A Component-Based Programming Model for Grid Web Services, XCAT. http://www.extreme.indiana.edu/xcat/index.html